**Article:**
Pellacani, Davide, Droop, Alastair P, Frame, Fiona M et al. (5 more authors) (2018) Phenotype-independent DNA methylation changes in prostate cancer. British journal of cancer. GG-2018-5107R. ISSN 1532-1827

https://doi.org/10.1038/s41416-018-0236-1

# Phenotype-independent DNA methylation changes in prostate cancer

**Running title: Prostate cancer specific DNA methylation**

Davide Pellacani[1,2*], Alastair P. Droop[1,3], Fiona M. Frame[1], Matthew S. Simms[4], Vincent M. Mann[1], Anne T. Collins[1], Connie J. Eaves[2] and Norman J. Maitland[1,5]

1) Cancer Research Unit, Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK.
2) Terry Fox Laboratory, BC Cancer Agency, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada
3) Leeds Institute for Data Analytics, Worsley Building, Clarendon Way, Leeds LS2 9NL, UK
4) Department of Urology, Castle Hill Hospital (Hull & East Yorkshire Hospitals NHS Trust), Cottingham HU16 5JQ, UK
5) Hull York Medical School, University of York, Heslington, York YO10 5DD, UK.

**Correspondence:**
Dr. Davide Pellacani

Terry Fox Laboratory,
BC Cancer Research Centre
675 West 10th Avenue
Vancouver, BC
V5Z 1L3, Canada

**Email:** dpellacani@bccrc.ca
**Phone:** 604-675-8120 Ext 7731
**FAX:** 604-877-0712

*Abstract*

**Background**: Human prostate cancers display numerous DNA methylation changes compared to normal tissue samples. However, definitive identification of features related to the cells' malignant status has been compromised by the predominance of cells with luminal features in prostate cancers.

**Methods**: We generated genome-wide DNA methylation profiles of cell subpopulations with basal or luminal features isolated from matched prostate cancer and normal tissue samples.

**Results**: Many frequent DNA methylation changes previously attributed to prostate cancers are here identified as differences between luminal and basal cells in both normal and cancer samples. We also identified changes unique to each of the two cancer subpopulations. Those specific to cancer luminal cells were associated with regulation of metabolic processes, cell proliferation and epithelial development. Within the prostate cancer TCGA dataset, these changes were able to distinguish not only cancers from normal samples, but also organ-confined cancers from those with extra-prostatic extensions. Using changes present in both basal and luminal cancer cells, we derived a new 17-CpG prostate cancer signature with high predictive power in the TCGA dataset.

**Conclusions**: This study demonstrates the importance of comparing phenotypically matched prostate cell populations from normal and cancer tissues to unmask biologically and clinically relevant DNA methylation changes.

## *Background*

Treatment-naïve prostate cancer (PCa) is characterized by an abnormal accumulation of proliferative cells with a molecular phenotype similar to the luminal cells present in the normal prostate[1,2]. However, PCa samples also contain a small population of tumour cells with basal features. These cells possess "cancer stem cell" features, appear to be treatment-resistant, and are proposed to serve as a reservoir for tumour recurrence after castration therapy[3-6]. DNA methylation of bulk PCa samples has been well studied[7] and aberrant methylation of promoter regions found to be a consistent feature[8], albeit with high variability both between patients and within single tumours[9]. Their frequency and presence in pre-malignant tissues support a strong selective pressure for DNA methylation changes during cancer development[7]. However, DNA methylation is dynamically regulated during tissue development and cell differentiation[10], and distinct cell types possess specific DNA methylation profiles within the same tissue[11-13]. Therefore, the luminal molecular features of bulk PCa samples, in contrast to the almost equal proportion of basal and luminal cells in normal prostate tissues, complicate the interpretation of datasets generated on whole tissue extracts, where changes associated to differences in cell types may mask the presence of malignancy-associated signatures.

Recent developments in tissue processing and the identification of surface markers suitable for the prospective isolation of viable basal and luminal cells from normal prostate tissues have enabled studies of their molecular and biological characteristics[14-17]. Use of this approach has revealed that many of the genes downregulated in normal luminal cells compared to basal cells are frequently hypermethylated in PCa[18]. This data implies a functional link between DNA

hypermethylation and the observed expansion of cells with a luminal phenotype in PCa. However, very little is known about the specific DNA methylation features of PCa cells with basal and luminal phenotypes in comparison to their normal counterparts. To address this issue, we generated genome-wide DNA methylation profiles of FACS-purified populations of cells with basal and luminal features from a series of freshly isolated patient-matched tumour and normal samples from individuals undergoing radical prostatectomy. Our results show that many DNA methylation changes frequently seen in PCa are characteristic differences between luminal and basal cells from both normal and cancer samples. From these datasets, we were also able to identify two sets of tumour-specific changes of potential clinical interest. One set consists of changes that are specific to PCa luminal cells; the other set are changes shared by both basal and luminal tumour but not normal prostate cells.

## *Methods*

### Tissue processing:

Prostate tissues were obtained from patients undergoing radical prostatectomy at Castle Hill Hospital (Cottingham, UK) with informed patient consent and approval from the NRES Committee Yorkshire & The Humber (LREC Number 07/H1304/121). Tissues were sampled immediately after surgery. For radical prostatectomies, three core needle biopsies were taken from four different sites (left base, left apex, right base, right apex) and were directed by previous pathology, imaging and palpation. Tissues were transported in RPMI-1640 with 5% FCS and 100U/ml antibiotic/antimitotic solution at 4ºC, and processed immediately upon arrival. PCa diagnosis was confirmed by histological examination of the whole prostate. Tissues were disaggregated as previously described[19], and all reagents were supplemented with 10 nM R1881 to better preserve the viability of luminal cells.

### Fluorescence activated cell sorting (FACS) and characterization of cell populations:

Single-cell suspensions were labelled with Lineage Cell Depletion Kit (human) and CD31 MicroBead Kit (Miltenyi Biotec) and $Lin^+/CD31^+$ cells depleted twice using MACS LS Columns (Miltenyi Biotec). $Lin^-/CD31^-$ cells were then labelled with EpCAM-APC, CD49f-FITC and CD24-PE (Miltenyi Biotec) and DAPI and $EpCAM^+/CD49f^+/CD24^-$ and $EpCAM^+/CD49f^-/CD24^+$ sorted at >95% purity using a MoFlo (Beckman Coulter) cell sorter. Sorted populations were characterized by immunofluorescence and qRT-PCR as previously described[18].

**Reduced Representation Bisulphite Sequencing (RRBS):**

DNA was extracted from FACS-sorted populations using phenol/chloroform extraction and ethanol precipitation. DNA was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific), and shipped to Zymo Research for RRBS analysis. Bisulphite conversion, library preparation, sequencing, and initial bioinformatics analyses were performed by Zymo Research following the Methyl-MiniSeq pipeline.

**Sequence data processing and methylation calls:**

Fastq files were trimmed using Trim Galore! v0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with the following parameters: --fastqc --illumina --paired --rrbs --non_directional. Trimmed sequences were aligned to the human genome (hg19 downloaded from UCSC, 08-Mar-2009 version) using bsmap v2.90[20] and the following parameters: -m 0 -x 1000 -n 1 -p 8 -S 1. The resulting bam files were sorted and indexed using samtools v0.1.19[21], and methylation and coverage calls for each CpG site calculated using the methratio.py script in the bsmap software (Supplementary Table 1). Methylation calls were then filtered for low (<3) and high (>99.95%) read coverage and merged in non-overlapping genomic bins of 100 bp using the methylKit package v0.99.2[22] within R v3.3.1 to increase comparability between samples. All subsequent analyses were carried out using only those genomic bins covered in all samples, with the exception of the results presented in Supplementary Fig. 2 and Supplementary Table 3 which were generated using single GpG information.

**Identification of differentially methylated regions (DMRs):**

DMRs were calculated using methylKit[22]; with all pairwise comparisons between the four cell populations carried out and similar populations from different donors defined as biological replicates. The patient of origin was used as a categorical covariate to account for the strong inter-donor variability seen. All p-values were generated using a logistic regression model and corrected for multiple testing using the SLIM method[23]. DMRs were defined as those genomic bins with q-values <0.05 and absolute methylation difference >10% in each pairwise comparison.

**Characterization of DMRs:**

All genomic features were downloaded from the UCSC Table browser (genome.ucsc.edu) for the hg19 genome. Gene models: "refGene" (RefSeq Genes), CpG Islands: "cpgIslandExt", Evolutionary conservation: "phastCons100way", DNase hypersensitivity sites (DHSs): "wgEncodeRegDnaseClusteredV3", transcription factor binding sites (TFBSs): "wgEncodeRegTfbsClusteredV3", repeats: "rmsk" (RepeatMasker). Overlaps and distances of DMRs to other genomic features were calculated using BEDtools v2.26.0[24], and significance of enrichments or depletions was calculated using custom R scripts. All p-values $<10^{-300}$ were approximated to $10^{-300}$ to avoid reaching the minimum value for a floating-point number ($2.2*10^{-308}$). Average conservation signals around DMRs were calculated using bwtool v1.0[25]. P-values were calculated using a bootstrapping approach comparing the average conservation of the distal DMRs with the average of an equal number of randomly selected, non-overlapping, distal genomic bins, 1000 times. Gene ontology (GO) analysis was performed using

GREAT v3.0[26], using all covered genomic bins as background and the default "Basal plus extension" association rules. Results were filtered to include only GO categories, with a Benjamini–Hochberg corrected (FDR) hypergeometric test p-value <0.05 and ≥3 genes with associated regions. K-means clustering of GO categories (biological processes only) was based on information similarity values calculated using the GOSim package within R v3.3.1. Promoters frequently altered in PCa were downloaded from the review by Massie et al., 2017[7]. Only promoters reported by ≥3 studies were considered frequently altered. Genome browser plots were generated using the package Sushi within R v3.3.1 and custom scripts.

**TCGA data analysis:**

Illumina Infinium HumanMethylation450 data generated within the The Cancer Genome Atlas (TCGA) consortium[27] was downloaded (pre-processed Level 3 data only) from the NCI Genomic Data Commons website using the provided GDC Data Transfer Tool (data downloaded on 7th Dec 2016). Clinical data was downloaded from firebrowse.org (8th Dec 2016). The presence of evident batch effects was excluded by visualizing the data on TCGA Batch Effects (http://bioinformatics.mdanderson.org/tcgambatch/). A data matrix containing the beta values for each sample was generated using custom scripts. Probes were mapped to hg19 using the positions officially reported by Illumina. Overlap of array probes with DMRs was carried out using BEDtools v2.26.0. Hierarchical clustering was based on Euclidean distances of unscaled beta-values. Logistic model training using least absolute shrinkage and selection operator (LASSO) regression was performed using the glmnet package within R v3.3.1 on a random selection of 70% of the samples. 200 lambda values

ranging from $e^{-7}$ to $e^{-2}$ were tested and 10-fold cross validation performed. The lambda with the minimum mean cross-validated error was selected and resulted in 17 probes with non-zero coefficients. The optimal model was then tested on the remaining 30% of samples and receiver operator curve and area under the curve (AUC) calculated using the ROCR package.

## *Results*

### Phenotypically defined prostate cells from patient-matched normal and PCa samples show donor-specific DNA methylation profiles

Matched tumour-directed (cancer) and contralateral (normal) core needle biopsies (1 or 2 per site) were obtained from 4 treatment-naïve prostate cancer patients undergoing radical prostatectomies. These samples were then enzymatically dissociated and labeled with antibodies against EpCAM, CD49f and CD24 to enable the prospective isolation of luminal (EpCAM+CD49f-CD24+) and basal (EpCAM+CD49f+CD24-) cells at >95% purity (Fig. 1A). EpCAM+CD49f+CD24- cells expressed higher levels of molecular markers associated with basal cells and lower levels of luminal markers compared to EpCAM+CD49f-CD24+ cells from the same biopsy, both at the mRNA and protein level (Supplementary Fig. 1A-B). For convenience, we named the paired subsets as follows: Cancer Luminal (CL) EpCAM+CD49f-CD24+ cells purified from tumour-directed biopsies; Cancer Basal (CB) EpCAM+CD49f+CD24- cells purified from tumour-directed biopsies; Normal Luminal (NL) EpCAM+CD49f-CD24+ cells from contralateral biopsies; Normal Basal (NB) EpCAM+CD49f+CD24- cells purified from contralateral biopsies. This yielded 4 CL and CB populations, and 3 matched NL and NB populations, as in one prostate the palpable tumour was extended to most of the prostate and it was not possible to obtain a contralateral "normal" tissue biopsy (Supplementary Fig. 1C). DNA obtained from each of these isolates was then subjected to Reduced Representation Bisulphite Sequencing (RRBS). On average, this generated information on the DNA methylation status of $>8.9 \times 10^6$ cytosines within CpG sites per sample (range $8 \times 10^6$ – $9.6 \times 10^6$, with an average coverage of 7.5 reads, Supplementary Table 1). The data was

182 processed as described in Methods, and binned into 100 bp genomic regions to maximize
183 the comparability between samples (932,905 bins covering $4.1x10^6$ CpGs in all samples).
184 Unsupervised hierarchical clustering of the top 1% most variable regions (bins) across all
185 samples showed clustering primarily according to the patient of origin, rather than the
186 subset analyzed (Fig. 1B). This indicates a high donor-determined variation in CpG
187 methylation, consistent with previous reports of similarly accrued data[28].
188
189 **Distinct DNA methylation profiles in basal and luminal cells**
190 We then calculated DMRs for all pairwise comparisons between the 4 sorted populations
191 (Fig. 1C, Supplementary Table 2). Among these, the comparison between CB and NB
192 cells (CB-NB comparison) produced the smallest number of DMRs. In contrast, a large
193 number of DMRs were seen when either normal or cancer luminal cells were compared
194 with either source of basal cells (i.e., NL-NB, NL-CB, CL-NB and CL-CB, Fig. 1D). Of
195 the DMRs revealed in these latter comparisons, ~2/3 were hypermethylated in luminal
196 cells, which correlates with the higher levels of DNMT3a seen in these cells[18]. We also
197 calculated differential methylation on single CpGs (prior the 100bp binning) with very
198 similar results (Supplementary Fig. 2 and Supplementary Table 3). Moreover, integration
199 of the DMRs identified in NL-NB proximal (±5 kb) to annotated transcriptional start sites
200 (TSSs) with RNA-seq data of similarly purified cells[15] showed the expected inverse
201 correlation (Supplementary Fig. 3A).
202 We also found an extensive overlap in the DMRs obtained from both the NL-NB and NL-
203 CB comparisons, and also from the CL-NB and CL-CB comparisons (Supplementary Fig.

3B-C). Accordingly, we focussed our subsequent analyses on comparisons of NL-NB and CL-CB, where cells from the same biopsy could be compared directly.

Characterization of the genomic features of the DMRs thus identified showed that >50% of them fell outside of CpG islands, shores or shelves (Fig. 1E), and >70% were >5 kb away from any annotated TSSs (Fig. 1F-G). These features were particularly pronounced (highly significant hypergeometric test) for the hypomethylated DMRs identified in the comparisons of NL-NB, CL-CB and CL-NL. Because hypermethylated and hypomethylated DMRs might be anticipated to differ in their genomic context, their impact on the biological properties of basal and luminal cells could also be different.

**Distal hypermethylated DMRs are enriched in enhancer features**

Given that most of the DMRs identified were outside CpG islands and far from TSSs, we asked whether they might affect distal regulatory elements (enhancers). We therefore examined three genomic characteristics of such elements: evolutionary conservation[29], open chromatin shown by hypersensitivity to DNase I[30], and presence of TFBSs[31]. Distal hypermethylated DMRs in each comparison were enriched for evolutionarily conserved sequences (Fig. 2A, bootstrapped p-value) and overlapped significantly with both DHSs and ChIP-seq-defined TFBSs (identified within the ENCODE project, Fig. 2B-C, hypergeometric test). Distal hypomethylated DMRs generally scored lower than the hypermethylated counterparts for each metric measured. DMRs hypomethylated in the CL-CB and CL-NL comparisons showed the weakest enrichments. However, all distal hypomethylated DMRs had high overlaps with genomic repetitive elements (Fig. 2D).

Specifically, LINE and LTR elements, but not SINE elements, were significantly enriched in the distal CL hypomethylated regions.

GO enrichment analysis (Fig. 2E, Supplementary Fig. 4) showed that hypermethylated DMRs in NL-NB were enriched for more than 500 terms, many of which were linked to prostate development or epithelial stem cell regulation; while hypomethylated DMRs in the same comparison were enriched for terms related to androgen receptor signalling and response to cytokines. In the CL-CB comparison, hypermethylated DMRs were also enriched for more than 500 terms, 311 of which were also identified in the NL-NB comparison, suggesting a high functional overlap in hypermethylated regions in luminal cells from both normal and cancer samples. In the CL-NL comparison, hypermethylated DMRs were enriched in terms related to cell adhesion, while hypomethylated DMRs were enriched in terms related to epithelial morphogenesis. These results indicate that several pathways fundamental to the establishment and maintenance of the normal prostate epithelium are altered in cancer cells with a luminal phenotype.

**Phenotype-specific DMRs are shared in normal and cancerous prostate tissues**

As suggested by the enriched GO analyses, we found a 28% overlap in all the DMRs identified from the NL-NB and the CL-CB comparisons (3852/13816, Fisher's exact test p−value < $10^{-300}$, Fig. 3A). Hierarchical clustering of all samples based on both sets of DMRs separated them by phenotype (Fig. 3B), reinforcing the presence of a strong phenotypic signature independent of disease state. These shared DMRs were enriched in features characteristic of enhancers (Supplementary Fig. 5A-D) and linked to GO terms

related to prostate development, regulation of epithelial stem cells and androgen receptor signalling (Supplementary Fig. 5E-F). Moreover, hypermethylated DMRs were highly enriched for TFBSs of *TP63*, *TP53* and *NF1*, and hypomethylated DMRs for *FOXA1*, *p65-NFkB* and *GATA3* (Fig. 3C), all well-known regulators of basal and luminal epithelial cells, respectively. Interestingly, 26 of the 168 genes described as frequently differentially methylated in PCa[7], showed hyper- or hypomethylated DMRs within 5 kb of their TSSs in both the NL-NB and CL-CB comparisons (Fig. 3D). These included the frequently hypermethylated genes, *GSTP1* and *CCDC8* (Fig. 3E-F).

In summary, these analyses identified a large set of phenotype-specific and disease-independent DMRs, both of which contained many binding sites for TFs with known regulatory roles in the normal prostate.

**CL hypermethylate PRC2 target sites and hypomethylate repetitive elements**

A second group of genes frequently hypermethylated in PCa were found hypermethylated in both the CL-CB and CL-NL comparisons (Fig. 4a), but not in the NL-NB comparison. These might be expected to reflect a PCa-specific methylation signature. DMRs identified in the CL-CB and CL-NL comparisons showed that many were shared (1472 DMRs, Fisher's exact test p−value < $10^{-300}$, Fig. 4B) with very few also different between NL and NB cells (106 DMRs). 65% of these CL-specific hypermethylated DMRs were distal to TSSs and were again highly enriched for enhancer features, but significantly depleted in repetitive elements (Supplementary Fig. 6A-E). These regions were associated with GO terms related to metabolic processes, cell proliferation and epithelial development (Fig. 4C) and showed a high enrichment of DNA sequences potentially

bound by EZH2 and SUZ12, two main members of the PRC2 complex (Supplementary Fig. 6F). On the other hand, distal hypomethylated DMRs were not enriched for any feature of putative regulatory regions, but significantly overlapped with LINE and LTR elements.

Since the CL subset represents the majority of the cells in untreated PCa samples, we hypothesized that aberrant methylation of these DMRs would be measurable even when whole tissue homogenates are analysed. We therefore interrogated the DNA methylation array dataset for PCa made available by the TCGA consortium, which consists of 50 PCa samples with matched normal counterparts, 452 additional PCa samples without normal counterparts, and 1 metastatic PCa sample[27]. 255 array probes overlap these 1472 DMRs. Hierarchical clustering of the 50 matched normal and PCa samples showed an almost perfect subdivision based on the malignancy status of the samples (TPR = 0.92, TNR = 0.92, Chi-squared test p-value = $2.4 \times 10^{-16}$, Fig. 4D). The same analysis carried out on all 553 samples produced similar results, with one cluster highly enriched in normal samples (Chi-squared test p-value = $1.7 \times 10^{-39}$, Supplementary Fig. 6G). This clustering also appeared to divide the PCa samples into two main groups, according to their differences from the normal samples. Exclusive analysis of the cancer samples confirmed this clustering pattern (Fig. 4E) and showed one cluster to be significantly enriched for samples with extra-prostatic extensions (pT3 or pT4 in TNM classification, Chi-squared test p-value < 0.005) in the absence of significant differences in Gleason score (Chi-squared test p-value >0.1).

Overall, these results indicate that phenotypic luminal PCa cells possess an aberrant methylation signature characterized by hypermethylation of putative regulatory

sequences involved in tissue development, and hypomethylation of LINEs and LTRs repetitive elements. This signature was also able to distinguish cancer samples from normal, and organ-confined from extraprostatic disease.

**Identification of PCa-specific, phenotype-independent DMRs**

Comparisons of the DMRs in the CL-NL and CB-NB pairs showed a small but significant overlap of both hyper- and hypomethylated DMRs in each (189 DMRs in total, Fig. 5A). These common DMRs were able to cluster all samples according to their disease state in a phenotype-independent manner (Supplementary Fig. 7A). Notably, they included DMRs close to many genes previously implicated in prostate cancer (e.g., *NEAT1*, *MTOR*, *RHCG*, *KCNC2*, *WT1*, *HOXC12*, *KMT2B*, Fig. 5B). To determine whether these DMRs would be altered in an independent dataset, we applied the same analysis to the TCGA dataset, where 66 array probes overlapped these 189 DMRs. Hierarchical clustering of the 50 matched normal and PCa samples produced a single cluster containing 46/50 normal samples and 10/50 PCa samples (TPR = 0.8, TNR = 0.92, Chi-squared test p-value = $1.8 \times 10^{-12}$, Fig. 5C). Application of the same analysis to all samples in the TCGA database produced similar results: one cluster was highly enriched in normal samples (TPR = 0.87, TNR = 0.74, Chi-squared test p-value = $8.3 \times 10^{-26}$, Supplementary Fig. 7B), indicating that at least some of these DMRs are frequently altered in PCa.

To select the probes most strongly associated with disease state (i.e., PCa vs normal), we trained a logistic model using LASSO regression on 70% of the TCGA samples and selected a 17-probe signature (Fig. 5D). We then tested this model on the

remaining 30% of the dataset. This resulted in an AUC of 0.92 (TPR = 0.9, TNR = 0.94, Fisher's exact test p−value = $2.82 \times 10^{-12}$ at the selected cut-off of 0.8, Fig. 5E-F, Supplementary Table 4). The 17-probe signature also included sequences proximal to several genes with recognized importance in PCa (e.g., *PLAGL1*/*HYMAI*, *HOXC12*, *KCNC2*), but was completely non-overlapping with other similar signatures recently developed for PCa[32-36].

***Discussion***

PCa is characterized by frequent aberrant DNA methylation of many genomic sites that may contain clinically relevant signatures[7,37]. The early establishment (presence in pre-neoplastic tissues) and high prevalence of these aberrant patterns is also suggestive of their direct involvement in PCA tumorigenesis. However, the normal prostate epithelium is composed of similar numbers of luminal and basal cells, whereas most treatment-naïve prostate cancers are largely composed of cells with many luminal features. This shift in favor of a transcriptional and epigenomic program of normal luminal cells might mask or complicate the identification of cancer-specific features in prostate cancer when bulk analyses are performed on this type of tumour.

We now report a detailed comparison of genome wide methylation profiles obtained separately from epithelial cells with luminal and basal phenotypes, isolated with a high purity from patient-matched normal and cancer biopsy samples. From comparative analyses of these profiles, we found a major proportion of the methylation differences between normal basal and luminal cells were conserved in their malignant counterparts. These affected many promoters frequently described as aberrantly methylated in bulk

PCa compared to normal tissues, consistent with the increased representation of cells with a luminal phenotype in PCa, in which a higher proportion of cells carrying a methylation signature of normal luminal cells might then be expected.

However, our study made it possible to identify, for the first time, regions specifically altered in the luminal fraction of PCa. The hypermethylated DMRs in this group were genes associated to genes involved in metabolic processes, cell proliferation and epithelial development, all functions clearly deregulated in prostate cancer, therefore potentally containing major cancer driver events. Furthermore, hypomethylated DMRs were highly enriched in repetitive elements, a feature also previously reported in many cancer types, where they have been thought to contribute to genomic instability and aberrant gene expression[38-40].

Importantly, this set of DMRs was able to discriminate not only normal and PCa samples in the TCGA dataset, but also PCa samples with or without extra-prostatic extensions, the former being indicative of highly aggressive, invasive cancers. Since this distinction was not evident from the Gleason grades of these tumours, the epigenetic data may reflect a an acquisition of specific aberrant epigenomic changes that herald disease progression[7,41-43]. Genomic regions consistently altered in both tumour phenotypes in the PCa samples analyzed also have potential clinical importance. Indeed, the new logistic model constructed from these regions makes use of only 17 probes to distinguish normal and PCa samples with similar specificity and sensitivity to previously developed, non-overlapping models[35,36], and may be useful in the context of the low mutagenic burdens seen in most hormone-naïve prostate cancers.

The results reported here show that many DNA methylation changes commonly associated with PCa cells are explained by a predominant luminal phenotype of the treatment-naïve PCa population, and are not cancer-specific nor are likely to contain driver events. Importantly however, we were able to identify two separate classes of PCa-specific DNA methylation changes: those specific to cancer luminal cells that can distinguish both normal from cancer samples and organ-confined cancers from those with extra-prostatic extensions; those changes common to basal and luminal cancer cells that are able to distinguish PCa efficiently from normal samples. These two novel sets of cancer-specific changes clearly demonstrate the potential of profiling normal and cancer cell subpopulations in identifying signatures that may contain previously unrecognized driver events in the development and progression of PCa.

## *Additional Information*

### Ethics approval and consent to participate

Prostate tissues were obtained from patients undergoing radical prostatectomy at Castle Hill Hospital (Cottingham, UK) with informed patient consent and approval from the NRES Committee Yorkshire & The Humber (LREC Number 07/H1304/121).

### Availability of data and materials

The methylation and coverage calls for all RRBS libraries generated are available from GEO [GSE107596]. For patients' privacy reasons, raw data (FASTQ and BAM files) for the RRBS libraries are not publicly available, but can be available from the corresponding author on request.

## Competing Interests

The authors declare no competing financial and non-financial interests.


## Funding

This work was supported by The Freemasons' Grand Charity (DP and NJM), Yorkshire Cancer Research program grant Y257PA (DP, NJM, FMF, APD, and ATC), and British Columbia Cancer Agency (Strategic Priorities Fund, DP and CJE).


## Authors' Contributions

DP and NJM designed the project. MSS and VMM procured the tissue samples. DP processed and sorted the tissue samples, and performed all other experiments. DP, FMF and ATC developed the tissue processing and sorting protocol. DP and APD conducted all bioinformatic analyses. DP, CJE and NJM wrote the manuscript. All authors contributed to the interpretation of the results and read and approved the manuscript.


## Acknowledgements

We thank the urology surgeons and patients from Castle Hill Hospital for kind donations of clinical prostate samples. We thank Artem Babaian, Rod Docking, Dr. Kieran O Neill, Hye-Jung E. Chun, Dr. Misha Bilenky, Dr. Alireza Heravi-Moussavi and Dr. Martin Hirst for the useful discussions regarding the bioinformatic analyses conducted in this study.

**References:**


1. Ananthanarayanan V, Deaton RJ, Yang XJ, Pins MR, Gann PH. Alteration of proliferation and apoptotic markers in normal and premalignant tissue associated with prostate cancer. BMC Cancer. 2006;6:73.

2. De Marzo AM, Meeker AK, Epstein JI, Coffey DS. Prostate stem cell compartments: expression of the cell cycle inhibitor p27Kip1 in normal, hyperplastic, and neoplastic cells. Am J Pathol. 1998 Sep;153(3):911–9.

3. Polson E, Lewis JL, Celik H, Mann VM, Stower MJ, Simms MS, et al. Monoallelic expression of TMPRSS2/ERG in prostate cancer stem cells. Nat Commun. 2013 Mar 27;4:1623.

4. Frame FM, Pellacani D, Collins AT, Simms MS, Mann VM, Jones G, et al. HDAC inhibitor confers radiosensitivity to prostate stem-like cells. Br J Cancer. 2013 Dec 10;109(12):3023–33.

5. Birnie R, Bryce SD, Roome C, Dussupt V, Droop A, Lang SH, et al. Gene expression profiling of human prostate cancer stem cells reveals a pro-inflammatory phenotype and the importance of extracellular matrix interactions. Genome Biol. 2008;9(5):R83.

6. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res. 2005 Dec 1;65(23):10946–51.

7. Massie CE, Mills IG, Lynch AG. The importance of DNA methylation in prostate cancer development. J Steroid Biochem Mol Biol. 2017 Feb;166:1–15.

8. Goering W, Kloth M, Schulz WA. DNA methylation changes in prostate cancer. Methods Mol Biol. 2012;863:47–66.

9. Aryee MJ, Liu W, Engelmann JC, Nuhn P, Gurel M, Haffner MC, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. Sci Transl Med. 2013 Jan 23;5(169):169ra10.

10. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, et al. Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. 2013 May 23;153(5):1149–63.

11. Farlik M, Halbritter F, Müller F, Choudry FA, Ebert P, Klughammer J, et al. DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. Cell Stem Cell. 2016 Dec 1;19(6):808–22.

12. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317–30.

13. Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp DJHF, Gakkhar S, et al. Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. Cell Rep. 2016 Nov 15;17(8):2060–74.

14. Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a cell of origin for human prostate cancer. Science. 2010 Jul 30;329(5991):568–71.

15. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, et al. Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. Nat Commun. 2016 Feb 29;7:10798.

16. Drost J, Karthaus WR, Gao D, Driehuis E, Sawyers CL, Chen Y, et al. Organoid culture systems for prostate epithelial and cancer tissue. Nat Protoc. 2016 Feb;11(2):347–58.

17. Karthaus WR, Iaquinta PJ, Drost J, Gracanin A, van Boxtel R, Wongvipat J, et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. 2014 Sep 25;159(1):163–75.

18. Pellacani D, Kestoras D, Droop A, Frame FM, Berry PA, Lawrence MG, et al. DNA hypermethylation in prostate cancer is a consequence of aberrant epithelial differentiation and hyperproliferation. Cell Death Differ. 2014 May;21(5):761–73.

19. Frame FM, Pellacani D, Collins AT, Maitland NJ. Harvesting Human Prostate Tissue Material and Culturing Primary Prostate Epithelial Cells. Methods Mol Biol. 2016;1443:181–201.

20. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics. 2009;10:232.

21. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

22. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012 Oct 3;13(10):R87.

23. Wang H-Q, Tuominen LK, Tsai C-J. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics. 2011 Jan 15;27(2):225–31.

24. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

25. Pohl A, Beato M. bwtool: a tool for bigWig files. Bioinformatics. 2014 Jun 1;30(11):1618–9.

26. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495–501.

27. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. 2015 Nov 5;163(4):1011–25.

28. Yu YP, Ding Y, Chen R, Liao SG, Ren B-G, Michalopoulos A, et al. Whole-genome methylation sequencing reveals distinct impact of differential methylations on gene transcription in prostate cancer. Am J Pathol. 2013 Dec;183(6):1960–70.

29. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009 May 7;459(7243):108–12.

30. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012 Aug 29;489(7414):75–82.

31. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3):144–54.

32. Zhao S, Geybels MS, Leonardson A, Rubicz R, Kolb S, Yan Q, et al. Epigenome-Wide Tumor DNA Methylation Profiling Identifies Novel Prognostic Biomarkers of Metastatic-Lethal Progression in Men Diagnosed with Clinically Localized Prostate Cancer. Clin Cancer Res. American Association for Cancer Research; 2017 Jan 1;23(1):311–9.

33. Geybels MS, Zhao S, Wong C-J, Bibikova M, Klotzle B, Wu M, et al. Epigenomic profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue. Prostate. 2015 Dec;75(16):1941–50.

34. Geybels MS, Wright JL, Bibikova M, Klotzle B, Fan J-B, Zhao S, et al. Epigenetic signature of Gleason score and prostate cancer recurrence after radical prostatectomy. Clin Epigenetics. BioMed Central; 2016;8(1):97.

35. Mundbjerg K, Chopra S, Alemozaffar M, Duymich C, Lakshminarasimhan R, Nichols PW, et al. Identifying aggressive prostate cancer foci using a DNA methylation classifier. Genome Biol. BioMed Central; 2017 Jan 12;18(1):3.

36. Tang Y, Jiang S, Gu Y, Li W, Mo Z, Huang Y, et al. Promoter DNA methylation analysis reveals a combined diagnosis of CpG-based biomarker for prostate cancer. Oncotarget. Impact Journals; 2017 Aug 29;8(35):58199–209.

37. Strand SH, Ørntoft TF, Sørensen KD. Prognostic DNA methylation markers for prostate cancer. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2014 Sep 18;15(9):16544–76.

519 38. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA
520 hypomethylation leads to elevated mutation rates. Nature. 1998 Sep
521 3;395(6697):89–93.

522 39. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors
523 promoted by DNA hypomethylation. Science. American Association for the
524 Advancement of Science; 2003 Apr 18;300(5618):455–5.

525 40. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human
526 cancer. Mob DNA. BioMed Central; 2016;7(1):24.

527 41. Ellinger J, Kahl P, Gathen Von Der J, Rogenhofer S, Heukamp LC, Gütgemann I,
528 et al. Global levels of histone modifications predict prostate cancer recurrence.
529 Prostate. 2010 Jan 1;70(1):61–9.

530 42. Angulo JC, Andrés G, Ashour N, Sánchez-Chapado M, López JI, Ropero S.
531 Development of Castration Resistant Prostate Cancer can be Predicted by a DNA
532 Hypermethylation Profile. J Urol. 2016 Mar;195(3):619–26.

533 43. McDonald OG, Li X, Saunders T, Tryggvadottir R, Mentch SJ, Warmoes MO, et
534 al. Epigenomic reprogramming during pancreatic cancer progression links anabolic
535 glucose metabolism to distant metastasis. Nat Genet. 2017 Mar;49(3):367–76.

536

537 ***Figure Legends***

538 **Fig. 1: Identification of DMRs between prostate cancer cell populations.** (**A**)

539 Representative FACS profiles of a cell suspension prepared from core needle biopsies of

540 a radical prostatectomy sample. (**B**) Heatmap showing scaled methylation values of the

541 top 1% most variable regions (100 bp bins) in the samples analyzed. Hierarchical

542 clustering is based on Euclidean distance of the unscaled values and complete linkage.

543 (**C**) Diagram showing all pairwise comparisons carried out. (**D**) Number of DMRs found

544 in each comparison. (**E**) Overlap of DMRs with CpG islands, shores (2 kb flanking

545 islands) or shelves (2 kb flanking shores). P-values from hypergeometric test against all

546 regions. E = enriched, D = depleted. (**F**) Distribution of distances of DMRs to the closest

547 TSS. Grey box indicates ±5 kb from a TSS. Purple lines: hypermethylated DMRs, orange

lines: hypomethylated DMRs, gray line: all regions. (**G**) Proportion of DMRs proximal or distal to TSSs. P-values from hypergeometric test against all regions. E = enriched, D = depleted.

**Fig. 2: Hypermethylated distal DMRs have features of enhancers.** (**A**) Average plots of evolutionary conservation scores of the distal DMRs in each set. Purple lines: hypermethylated DMRs; orange lines: hypomethylated DMRs, gray line: all regions. P-values are from bootstrapping analysis. (**B**) Proportion of distal DMRs overlapping with DHSs (identified by ENCODE). P-values from hypergeometric test against all regions. E = enriched, D = depleted. (**C**) Overlap of distal DMRs with ChIP-seq derived TFBSs (identified by ENCODE). P-values are from hypergeometric tests against all regions. E = enriched, D = depleted. (**D**) Overlap of each set of distal DMRs with repetitive elements (UCSC repeatMask), SINEs, LINEs and LTRs. P-values from hypergeometric tests against all regions. E = enriched, D = depleted. (**E**) Number of GO terms enriched by each set of DMRs. GO terms identified using GREAT (FDR<0.05 and at least 3 genes in the set).

**Fig. 3: Shared phenotype-specific DMRs.** (**A**) Overlap between the DMRs identified in the NL-NB and CL-CB comparisons. P-values derived from Fisher's exact test. (**B**) Heatmap showing scaled methylation values of the DMRs identified in the NL-NB (left) or CL-CB (right) comparisons. Hierarchical clustering is based on Euclidean distances of the unscaled values and complete linkage. (**C**) TFBSs enriched in the hypermethylated (purple) or hypomethylated (orange) DMRs common between the NL-NB and CL-CB

comparisons. Left panel: analysis performed using HOMER findMotifs, p-values from binomial test. Right panel: enrichment of ENCODE defined TFBSs, p-values from hypergeometric test against all regions. (**D**) Frequently hyper- or hypomethylated genes in PCa[7] that were also hypermethylated (purple) or hypomethylated (orange) in the NL-NB and CL-CB comparisons. (**E-F**) Genome browser plots of the promoter regions of GSTP1 (**E**) and CCDC8 (**F**). Grey squares are the bins analyzed. Lines and shaded areas represent mean ±SEM of each category (NB=light blue, NL=light red, CB=dark blue, CL=dark red). DMRs are shown on top: hypermethylated=purple, hypomethylated=orange.

**Fig. 4: Aberrant methylation in CL.** (A) Frequently hyper- or hypomethylated genes in PCa[7] that are also hypermethylated (purple) or hypomethylated (orange) in the CL-CB and CL-NL comparisons. (B) Overlap between the DMRs identified in the CL-CB and CL-NL comparisons. P-values derived from Fisher's exact test**.** (C) Clustering of the gene ontologies (biological process) enriched in DMRs common between the CL-CB and CL-NL comparisons based on information similarity. Each circle shows an individual GO term enriched in regions hypermethylated (purple), hypomethylated (orange) or both (green), the size of the circles is proportional to the enrichment p-value. The 2 main clusters of GO terms determined by k-means are highlighted (light blue and pink), and named after the most frequent terms. (D) Heatmap showing scaled methylation values (β-values) of probes overlapping the DMRs common to the CL-CB and CL-NL comparisons in the PCa samples (magenta) and matched normal samples (green) within the TCGA dataset. Hierarchical clustering based on Euclidean distances of the unscaled values and

complete linkage. The dark green and gray clusters were generated by cutting the tree at the first bifurcation. (E) Heatmap showing scaled methylation values (β-values) of probes overlapping the DMRs common to the CL-CB and CL-NL comparisons in the PCa samples (matched normal samples not included) of the TCGA dataset. Hierarchical clustering based on Euclidean distance of the unscaled values and complete linkage. The dark green and gray clusters are generated by cutting the tree at the first bifurcation.

**Fig. 5: PCa-specific DMRs shared between CB and CL.** (**A**) Overlap between the DMRs identified in the CL-NL and CB-NB comparisons. P-values derived from Fisher's exact test. (**B**) Genome browser views of KCNC2 promoter (top) and RHCG exon 2 (bottom). Grey squares are the bins analyzed. Lines and shaded areas represent mean ±SEM of each category (NB=light blue, NL=light red, CB=dark blue, CL=dark red). DMRs are shown on top: hypermethylated=purple, hypomethylated=orange. (**C**) Heatmap showing scaled methylation values of probes overlapping the DMRs common between CL-CB and CB-NB in the matched normal and cancer samples within the TCGA dataset. Hierarchical clustering based on Euclidean distances of the unscaled values and complete linkage. The dark green and gray clusters were generated by cutting the tree at the first 2 bifurcations. (**D**) Selection of a 17-probe signature distinguishing normal and PCa samples applying LASSO regression on a logistic model of the training dataset (70% of the TCGA samples). Lines show the changes in coefficients in relation to different lambdas. The vertical dashed line shows the optimal lambda identified using cross-validation. (**E**) Receiver-operating characteristic curve generated by applying the optimal logistic model to the test dataset (30% of the TCGA samples). (**F**) Heatmap showing

scaled methylation values of the 17-probe signature in the test dataset (30% of the TCGA samples). The bar plot on the left side shows the final coefficients for each probe in the model, and the bar plot on top shows the logistic probability generated by for each sample (Green: normal samples, magenta: cancer samples).

Figure 1

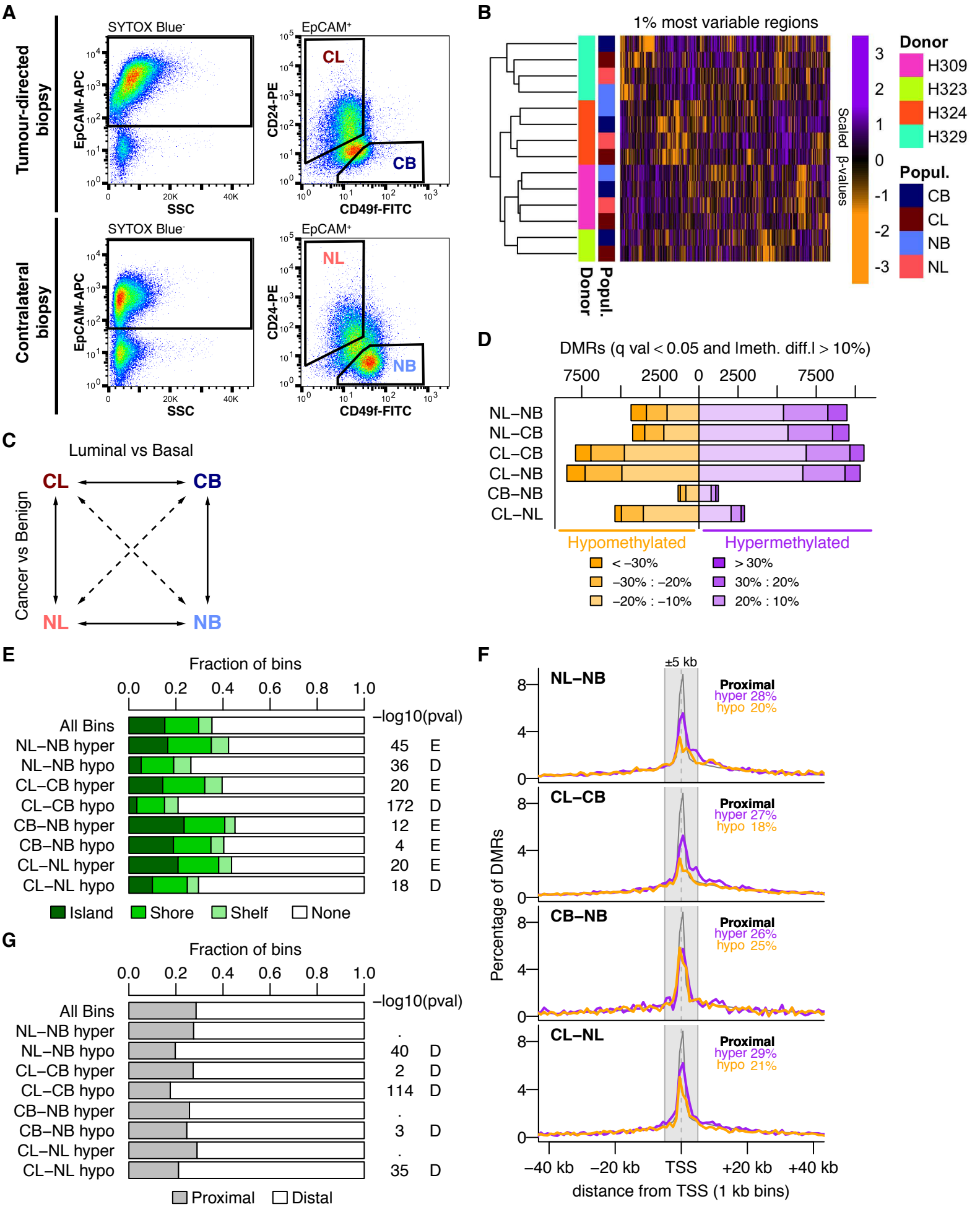

A
Tumour-directed biopsy
Contralateral biopsy
SYTOX Blue-
EpCAM+
EpCAM-APC
CD24-PE
SSC
CD49f-FITC
CL
CB
NL
NB
B
1% most variable regions
Donor
H309
H323
H324
H329
Popul.
CB
CL
NB
NL
Scaled β-values
C
Luminal vs Basal
Cancer vs Benign
D
DMRs (q val < 0.05 and |meth. diff.| > 10%)
NL–NB
NL–CB
CL–CB
CL–NB
CB–NB
CL–NL
Hypomethylated
Hypermethylated
< –30%
–30% : –20%
–20% : –10%
> 30%
30% : 20%
20% : 10%
E
Fraction of bins
–log10(pval)
All Bins
NL–NB hyper
NL–NB hypo
CL–CB hyper
CL–CB hypo
CB–NB hyper
CB–NB hypo
CL–NL hyper
CL–NL hypo
Island
Shore
Shelf
None
F
±5 kb
Proximal
hyper 28%
hypo 20%
hyper 27%
hypo 18%
hyper 26%
hypo 25%
hyper 29%
hypo 21%
Percentage of DMRs
–40 kb
–20 kb
TSS
+20 kb
+40 kb
distance from TSS (1 kb bins)
G
Proximal
Distal

**Figure 2**

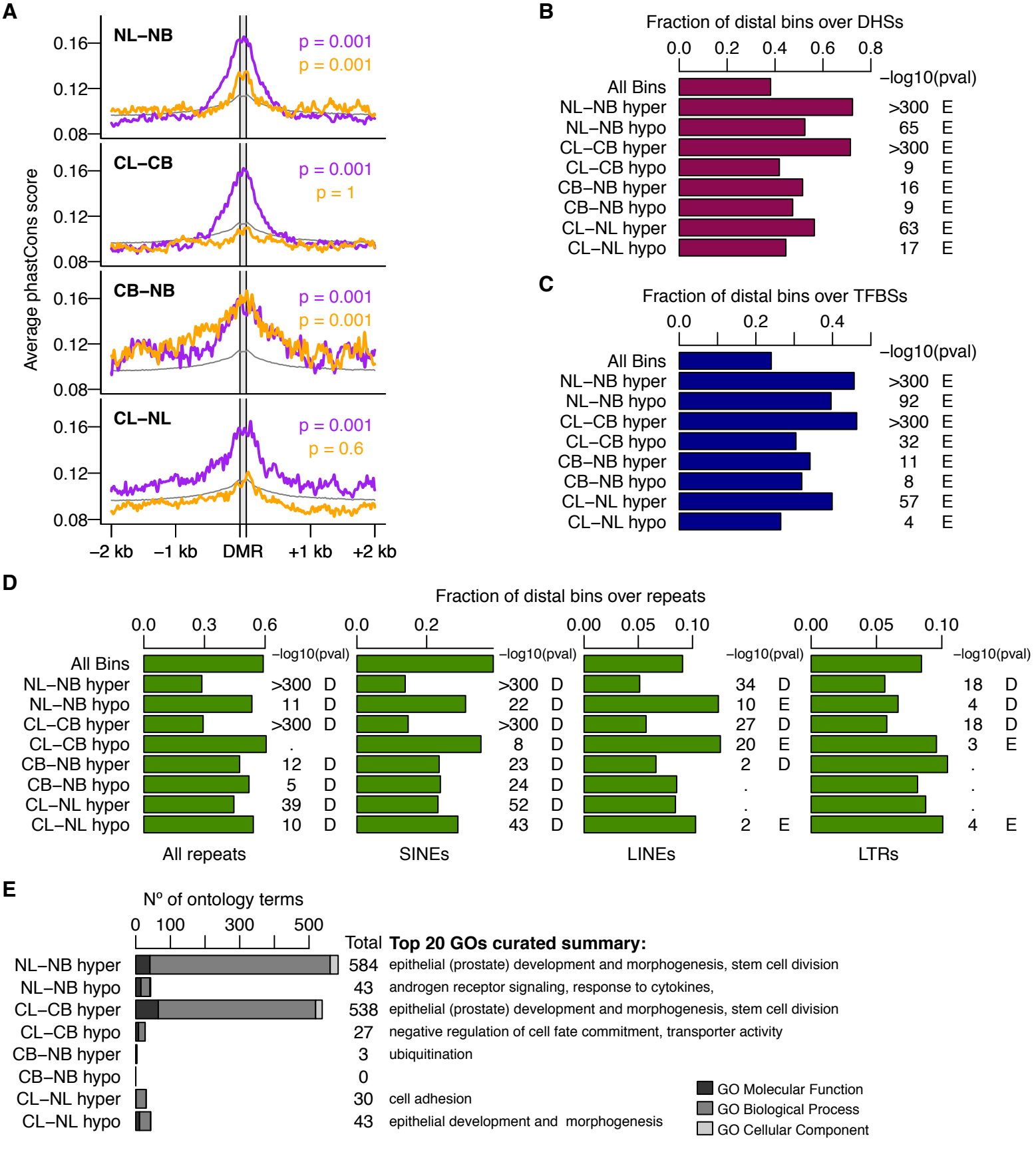

A
NL–NB
p = 0.001
p = 0.001
CL–CB
p = 0.001
p = 1
CB–NB
p = 0.001
p = 0.001
CL–NL
p = 0.001
p = 0.6
Average phastCons score
0.16
0.12
0.08
–2 kb
–1 kb
DMR
+1 kb
+2 kb
B
Fraction of distal bins over DHSs
0.0 0.2 0.4 0.6 0.8
–log10(pval)
All Bins
NL–NB hyper >300 E
NL–NB hypo 65 E
CL–CB hyper >300 E
CL–CB hypo 9 E
CB–NB hyper 16 E
CB–NB hypo 9 E
CL–NL hyper 63 E
CL–NL hypo 17 E
C
Fraction of distal bins over TFBSs
0.0 0.2 0.4
–log10(pval)
All Bins
NL–NB hyper >300 E
NL–NB hypo 92 E
CL–CB hyper >300 E
CL–CB hypo 32 E
CB–NB hyper 11 E
CB–NB hypo 8 E
CL–NL hyper 57 E
CL–NL hypo 4 E
D
Fraction of distal bins over repeats
All repeats
SINEs
LINEs
LTRs
–log10(pval)
All Bins
NL–NB hyper >300 D >300 D 34 D 18 D
NL–NB hypo 11 D 22 D 10 E 4 D
CL–CB hyper >300 D >300 D 27 D 18 D
CL–CB hypo . 8 D 20 E 3 E
CB–NB hyper 12 D 23 D 2 D .
CB–NB hypo 5 D 24 D . .
CL–NL hyper 39 D 52 D . .
CL–NL hypo 10 D 43 D 2 E 4 E
E
Nº of ontology terms
0 100 300 500
Total
Top 20 GOs curated summary:
NL–NB hyper 584 epithelial (prostate) development and morphogenesis, stem cell division
NL–NB hypo 43 androgen receptor signaling, response to cytokines,
CL–CB hyper 538 epithelial (prostate) development and morphogenesis, stem cell division
CL–CB hypo 27 negative regulation of cell fate commitment, transporter activity
CB–NB hyper 3 ubiquitination
CB–NB hypo 0
CL–NL hyper 30 cell adhesion
CL–NL hypo 43 epithelial development and morphogenesis
GO Molecular Function
GO Biological Process
GO Cellular Component

**Figure 3**

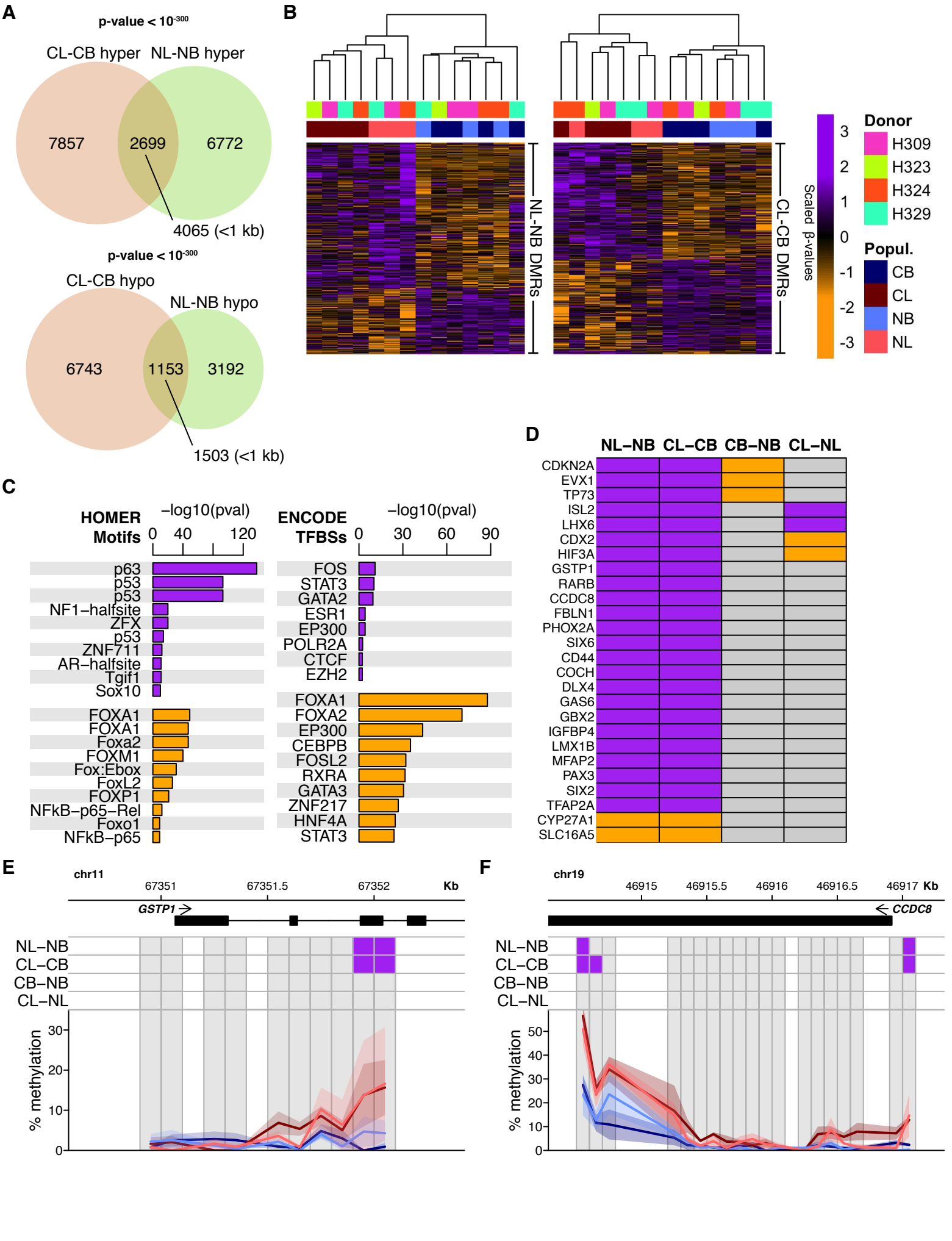

A
p-value < 10-300
CL-CB hyper
NL-NB hyper
7857
2699
6772
4065 (<1 kb)
p-value < 10-300
CL-CB hypo
NL-NB hypo
6743
1153
3192
1503 (<1 kb)
B
NL-NB DMRs
CL-CB DMRs
Scaled β-values
3
2
1
0
-1
-2
-3
Donor
H309
H323
H324
H329
Popul.
CB
CL
NB
NL
C
HOMER Motifs
−log10(pval)
0 40 120
p63
p53
p53
NF1−halfsite
ZFX
p53
ZNF711
AR−halfsite
Tgif1
Sox10
FOXA1
FOXA1
Foxa2
FOXM1
Fox:Ebox
FoxL2
FOXP1
NFkB−p65−Rel
Foxo1
NFkB−p65
ENCODE TFBSs
−log10(pval)
0 30 60 90
FOS
STAT3
GATA2
ESR1
EP300
POLR2A
CTCF
EZH2
FOXA1
FOXA2
EP300
CEBPB
FOSL2
RXRA
GATA3
ZNF217
HNF4A
STAT3
D
NL−NB
CL−CB
CB−NB
CL−NL
CDKN2A
EVX1
TP73
ISL2
LHX6
CDX2
HIF3A
GSTP1
RARB
CCDC8
FBLN1
PHOX2A
SIX6
CD44
COCH
DLX4
GAS6
GBX2
IGFBP4
LMX1B
MFAP2
PAX3
SIX2
TFAP2A
CYP27A1
SLC16A5
E
chr11
67351
67351.5
67352
Kb
GSTP1
NL−NB
CL−CB
CB−NB
CL−NL
% methylation
30
20
10
0
F
chr19
46915
46915.5
46916
46916.5
46917 Kb
CCDC8
NL−NB
CL−CB
CB−NB
CL−NL
% methylation
50
40
30
20
10
0

**Figure 4**

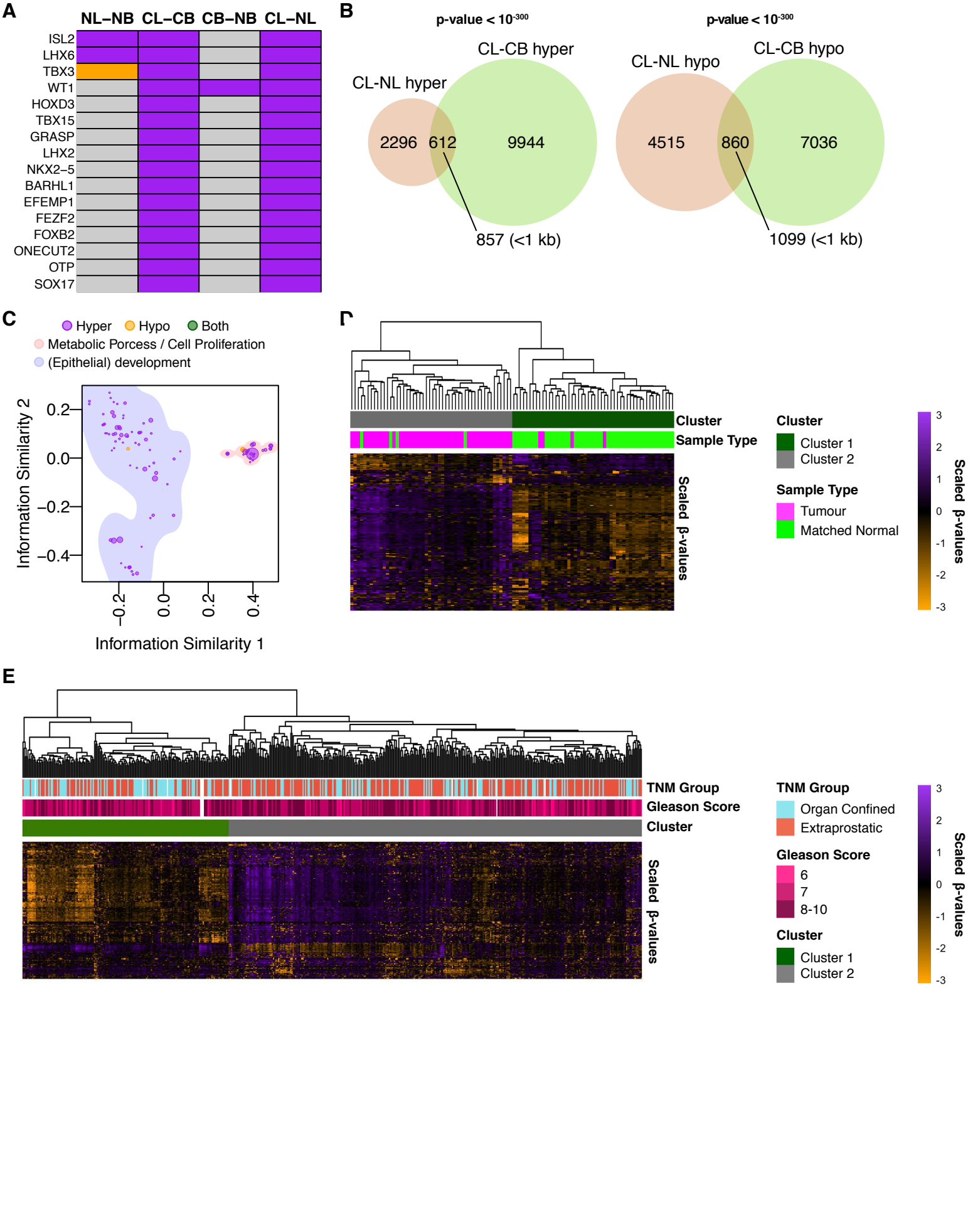
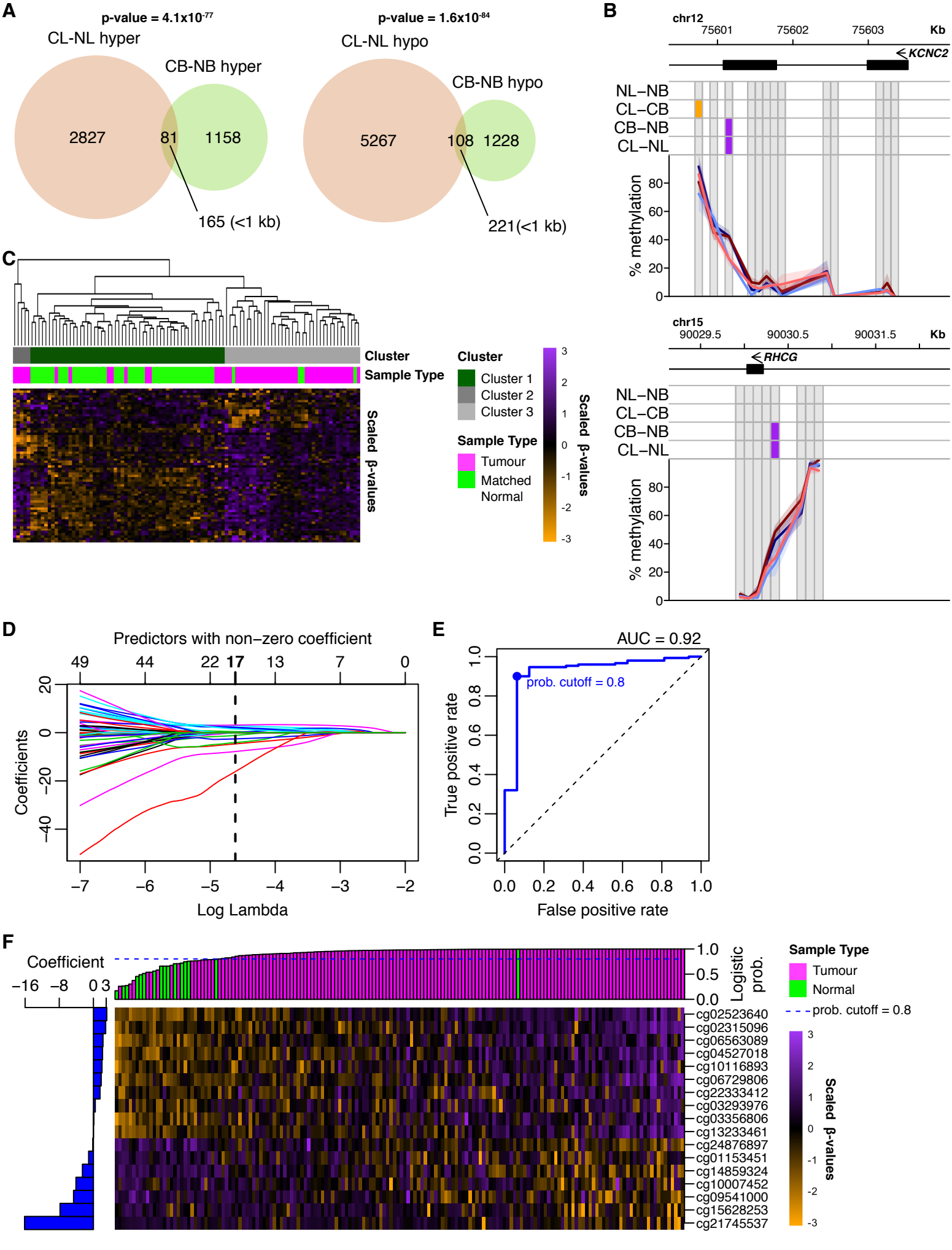

A
NL–NB
CL–CB
CB–NB
CL–NL
ISL2
LHX6
TBX3
WT1
HOXD3
TBX15
GRASP
LHX2
NKX2–5
BARHL1
EFEMP1
FEZF2
FOXB2
ONECUT2
OTP
SOX17
B
p-value < 10-300
CL-CB hyper
CL-NL hyper
2296
612
9944
857 (<1 kb)
p-value < 10-300
CL-NL hypo
CL-CB hypo
4515
860
7036
1099 (<1 kb)
C
Hyper
Hypo
Both
Metabolic Porcess / Cell Proliferation
(Epithelial) development
Information Similarity 2
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
0.4
Information Similarity 1
D
Cluster
Sample Type
Scaled β-values
Cluster
Cluster 1
Cluster 2
Sample Type
Tumour
Matched Normal
3
2
1
0
-1
-2
-3
Scaled β-values
E
TNM Group
Gleason Score
Cluster
Scaled β-values
TNM Group
Organ Confined
Extraprostatic
Gleason Score
6
7
8-10
Cluster
Cluster 1
Cluster 2
3
2
1
0
-1
-2
-3
Scaled β-values

**Figure 5**


A
p-value = 4.1x10-77
CL-NL hyper
CB-NB hyper
2827
81
1158
165 (<1 kb)
p-value = 1.6x10-84
CL-NL hypo
CB-NB hypo
5267
108
1228
221(<1 kb)
B
chr12
75601
75602
75603
Kb
KCNC2
NL–NB
CL–CB
CB–NB
CL–NL
% methylation
chr15
90029.5
90030.5
90031.5
Kb
RHCG
C
Cluster
Sample Type
Scaled β-values
Cluster 1
Cluster 2
Cluster 3
Tumour
Matched Normal
D
Predictors with non–zero coefficient
Coefficients
Log Lambda
E
AUC = 0.92
prob. cutoff = 0.8
True positive rate
False positive rate
F
Coefficient
Logistic prob.
Sample Type
Tumour
Normal
prob. cutoff = 0.8
Scaled β-values
cg02523640
cg02315096
cg06563089
cg04527018
cg10116893
cg06729806
cg22333412
cg03293976
cg03356806
cg13233461
cg24876897
cg01153451
cg14859324
cg10007452
cg09541000
cg15628253
cg21745537